Recurrence analysis techniques for non-stationary and non-linear data

Philip Kostov
Department of Agricultural and Food Economics, Queen's University Belfast, UK

John Lingard
Department of Agricultural Economics and Food Marketing, University of Newcastle, UK

Acknowledgements

This research was undertaken with support from the European Union's Phare ACE Programme. The content of the publication is the sole responsibility of the authors and it in no way represents the views of the Commission or its services.

Abstract

When analysing food consumption data a number of problems arise when one departs from the comparative statics of conventional demand theory. Two of these properties, non-linearity and non-stationarity present a major challenge for econometric modelling. A new method for time series analysis, namely recurrence analysis, is outlined which allows for robust analysis of data that can not be satisfactorily handled with established econometric methods. The method is explained and applied to specific food consumption data. General implications for empirical modelling of similar data are inferred.

JEL classification: C22, C40

Non-technical summary

Not available

1. Introduction

Food consumption has been a focus of attention for economists. It is well known that consumer choice can no longer be considered in the restricted framework of prices and incomes as conventional economic theory postulates. Numerous new determinants of food demand, such as preferences, cultural characteristics and enviromental issues are also important. The inclusion of these factors and influences in quantitative models of consumption is however difficult. They are mainly unobserved variables and whatever measures or proxies are used for them, there is a danger of misinterpretation.

Conventional consumption theory uses consumer preferences to explain consumer choices. This is however in the framework of comparative statics where preferences are regarded as fixed. When a more flexible dynamic approach is adopted, preferences can no longer be considered fixed. They change and indeed can be created. Take as an example organic or non-genetically modified food. Who had thought about this several years ago and had established preferences towards non-GM food? Preference changes are considered exogenous within conventional consumption theory - their change alters the nature of demand. When the underlying determinants of consumer choice are changeable, this implies that the relationships that econometricians estimate from time series data will themselves be variable. The mechanism that generates the economic data can be variable as the Austrian subjectivists suggested many years ago. Additionally the model of consumption or consumer choice even after imposing appropriate simplifications for the sake of tractabilty "... is still ... a non-linear program..." (Lancaster, 1966).

The conclusion is that any attempt to include variables other than prices and incomes in models of consumer behaviour, will inevitably introduce non-linear and time-varying relationships. With regard to standard econometric terminology, the additional implications can be defined as non-linearity and non-stationarity. The non-linear character of some economic time series poses problems for analysis. Often econometricians utilise linear models, relying upon finite Taylor series approximations hoping that they can capture the non-linear effect via appropriate representations of the irregular terms. The non-linearity, however, can be expressed in different forms, and, in some cases, there may not exist any appropriate linear approximations (Urbach, 2000). Ideally one would like to estimate the invariant characteristics of the unknown dynamic system that generated the time series. Unfortunately methods for doing so require long and noise-free series, which is normally not the case with the data available. Moreover, most of the available methods assume data stationarity. Therefore we need to use different tools for assessing data and extracting information when the data is non-stationary.

2. Data

We illustrate the above problem with analysis of data showing its implications for quantitative modelling in general. The data we use are seasonally and shock adjusted product indices of subsistence presented in Kostov (2001). These are presented graphically in appendix 1. They measure the degree of self-sufficiency of households showing the share of own production in household consumption for different products in Bulgaria. These indices are first calculated from the break-down of monthly household budgets data and represent a ratio of that part of household production that is left for own consumption to total household consumption. The formula for calculating the subsistence indices (SI) is:

SI = (household production - sold quantities) / balance sum.

Three products have been selected for the present study - beef, milk and veal. Intutively one would expect these cattle-based products to be related to each other and be part of the same system. Another point of interest is the detected non-linearity and non-stationarity in these series (Kostov, 2001).

The data is preliminarily adjusted for seasonality and for structural breaks, using conditionally Gaussian univariate unobserved components models estimated in state space form. Seasonal adjustment is necessary to rule out the seasonal effects, which in case of monthly agricultural data are considerable, and combined with the effect of stocks drive the subsistence indices away from the desirable interval (0,1) and make interpretation difficult. Model based seasonal adjustment has been preferred to the automatic and semi-automatic ARIMA-based methods, because of the possibility of structural working directly with non-differenced data, which in the case of numerous structural breaks produces superior results by utilising the information contained in the data in levels. Harvey and Koopman (2000) provide detailed discussion of the advantages of unobserved components models which also apply to the seasonal case. The preliminary adjustment has been carried out using the following model:

$$y_t = \mu_t + \gamma_t + x_t'b + \varepsilon_t$$
$$\mu_{t+1} = \mu_t + \beta_t + \eta_t$$
$$\beta_{t+1} = \beta_t + \xi_t$$
$$\gamma_t = (1, 0')\gamma_t^{\times}$$
$$\gamma_{t+1}^{\times} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{12} \end{pmatrix} = \rho\gamma_t^{\times} + \omega_t$$
with ε_t ~ NID(0, σ_ε^2), η_t ~ NID(0, σ_η^2), ξ_t ~ NID(0, σ_ξ^2) and
$$\omega_t = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_{12} \end{pmatrix}_t \sim \mathrm{NID}\left\{0, \sigma_\omega^2 (1-\rho^2)\frac{12 * \mathrm{I} - ii'}{11}\right\},$$
$$\gamma_1^{\times} \sim \mathrm{N}\left\{0, \sigma_\omega^2 \frac{12 * \mathrm{I} - ii'}{11}\right\},$$
where I is an 12x12 identity matrix and i is a 12x1 vector of ones.

This is a very simple model that represents the data series in terms of trend $\boldsymbol{\mu}$, seasonal $\boldsymbol{\gamma}$ regression effects **x** and irregular component $\boldsymbol{\varepsilon}$. The only regression effects used were dummies to account for the structural breaks. Outliers have not been adjusted. The adjusted data represents a sum of the trend and the irregular components. All the above components are variable and the trend component incorporates local slope $\boldsymbol{\beta}$. One can see from the data that there are considerable cyclical effects at non-seasonal frequencies which however, have not been modelled, since the primary aim was to account for the seasonal effects and structural breaks. The used seasonal representation is an extension to the time invariant form (Harvey et al., 1998) of the Harrison and Stevens (1976) seasonal representation that allows for both stationary and non-stationary seasonal components. The parametric bootstrap score test, suggested in Koopman et al. (1999) was implemented to specify which seasonal components are stationary. For completeness we note that while the seasonal unit root tests of Beaulieu and Miron (1993) found 5 unit roots at the seasonal frequencies for the beef series and 12 for the other two series, in the above model context only two of the seasonal components in the beef series, three in the case of milk and six for the veal series were non-stationary. This is of course due to the well known fact that inference in seasonal models depends crucially on the form of model adopted. We have paid much attention to the preliminary data manipulation, because correctly specifying and distinguishing stationary and non-stationary seasonal components is crucial for the validity of the following analysis. The properties of the deseasonalised data would critically depend on the form of the model used. In this case we have used the preliminary data manipulation only to estimate data series with known properties for applying recurrence analysis techniques. A proper modelling strategy would be to explicitly model other components such as cyclical components and only after fully specifying the model to subtract the seasonal effects. The latter would however distract our attention from the recurrence analysis methods which are the primary subject of this paper.

We would not expect any further structural breaks in these series. There is however concern that because the data comes from Bulgaria's unstable transitional economy, the models may not achieve appropriate adjustments to the non-linear shocks. The data used in this study covers the period from January 1989 to April 1996. Adjusted data after April 1996 shows sensitivity to the choice of adjustment method. Recurrence analysis has been applied to these series and the fundamental properties are then described.

3. Method

Recurrence Quantification Analysis (RQA) is based on the more qualitative recurrence plot tool introduced by Eckmann et al.(1987). The recurrence plot is defined in terms of the distance matrix between the rows of an embedded matrix of the scalar time series at a fixed lag. The sequence of vectors $\{x_k\}$ are embedded[1] in $\Re^n$ forming the sequence of points $\{y_i\}$[2]. We can define a function on an n x n array in the following way: darken the (i,j)th element of the array if $||y_i - y_j|| \leq r$, for a specified number $r > 0$, where $||.||$ is the (Euclidean) norm.[3] The result of the above is the recurrence plot. It is symmetric ($||y_i-y_j||=||y_j-y_i||$) with a darkened main diagonal ($||y_i-y_i|| \leq r$ for any choice of i and r). Zbilut and Webber (1992) develop a number of quantitative measures characterising the recurrence plot. Using these measures instead of the original recurrence plot allows quantitative analysis of the series and can be used to verify the results from visual inspection of the recurrence plots. It has been demonstrated that this method is robust with regard to stationarity constraints, number of observations and choice of embedding dimension (Webber and Zbilut, 1994). The measures are as follows:

<u>Recurrence</u> - percentage of the recurrent points in the total plot. It represents the share of recurrent pairs in all possible pairs of points.

<u>Determinism</u> - percentage of the recurrence points that are "ordered" in sequences forming line structures in the distance matrix parallel to the main diagonal. While recurrence may in principle occur by chance, this characteristic measures points that are close not only in embedding space but also in time. Eckmann et al. (1987) provide discussion on the relations of these diagonal lines to the Lyapunov spectra.
<u>Entropy</u> - an approximation of the Shanon entropy, calculated from the line structures of the recurrence plot.

---

[1] Strictly speaking the embedding may be constructed in terms of an n-dimensional manifold. We lose no generality by representing it in n-dimensional space.

[2] Using the fixed lag, usually referred to as a delay.

[3] It is possible to use other norms.

Maximum line - this is the length (consecutive points) of the longest recurrent line in the plot. This is an important characteristic of the recurrence plot, which is invariant with regard to its parameters.

Trend - calculated as an OLS coefficient in the regression of time, expressed in terms of distance from the main diagonal and the amount of recurrence. It represents the rate of change of the recurrence when moving forward in time.
The recurrence plot can be regarded as a general "autocorrelation function" of the series, because it relates its values at different times. Unlike the autocorrelation function, it takes into consideration the relative and also the absolute time, allowing analysis of non-stationary data.

4. Parameters of the analysis

We have chosen a delay parameter equal to one, which works well in practice for discrete time series. Nevertheless we need to determine the appropriate embedding dimension for carrying out the RQA. It is often (Webber and Zbilut, 1994) recommended to choose a high value for the embedding dimension to ensure real embedding. It is however not desirable to overestimate it because a greater embedding dimension will enhance the differences with the true system (Urbach, 2000; Hegger et al., 1999). We use the mutual information as a criterion for finding an appropriate embedding. The minimum of the mutual information is obtained at an embedding dimension of eight for beef and thirteen for veal. For milk the minimum is obtained at dimension value of 1, the second smallest value being at dimension value of seven. The result for milk is a reflection of the result of Kostov (2001) showing linear effects in the index for milk. It can be shown that the properties of the recurrence plot for linear and low-dimensional processes can be obtained without using embedding.

In the further analysis we apply embedding dimensions for the series, as determined by the minimum of the mutual information.

The algorithm used to calculate the mutual information, given by Hegger et al. (1999) is only valid asymptotically and does not employ finite sample corrections. In order to check the validity of the proposed embedding dimensions, we calculate the RQA statistics for embedding dimensions from seven to thirteen and different values of the radius for the original series, differenced series and the shuffled versions. The maximum recurrent line is used to check for validity of the embedding. Its value should not depend on the choice of recurrence plot parameters, in this case the radius. Therefore any change in the estimated length of the maximum recurrent line would mean invalid embedding. This check confirms that the embeddings based on the asymptotical minimum of the mutual information are valid. This is not the case with the differenced series suggesting that their appropriate embedding dimension is much higher.

5. Results

In appendix 2 we present a number of recurrence plots, realised at values of the radius approximately equal to the mean distance in the embedding space. The heterogeneity of these plots demonstrates changing through time relationships. Quantitatively this variability can be captured by the trend statistic. It shows the change in recurrence throughout time and can be used as a measure of stationarity when close to zero and non-stationarity when away from zero. If the recurrence rate changes when moving through time, then the properties of the time series change, showing time variable relationships. All the indices show non-stationarity.

The differenced series also look non-stationary which can be proved by the estimated trend statistic. We have to remember that the embedding dimensions used for these series are the same as in the levels. This may therefore be a spurious result, because of the inappropriate choice of dimension which is too low to ensure embedding. Since the model used to adjust the series includes a variable slope component (ß), we know that the differenced series in this case are indeed non-stationary. Nevertheless the assumption that economic data is stationary after differencing is so common in econometrics, that in the case of any other data that has not been subjected to preliminary manipulation, one would be tempted to find appropriate embedding to test for this assumption.

We check the variability of the maximum recurrent line for a number of high values of the embedding dimension for the differenced series. Table 1 presents some estimates of the maximum recurrent line according to the radius parameter. The results show variability in the maximum line parameter - for the higher dimensions this variability is rather small. Hence the impossibility to find embedding is probably due to noise in the data which causes the variability in the results. Due to the much smaller magnitude of the differenced compared to the original series, this noise has significantly greater impact on the latter. This casts doubt on the usefulness of the ARIMA methodology.Comparing the results for determinism and entropy of the original series and their shuffled version helps to detect deterministic time order. In all cases, these statistics change considerably between ordered and shuffled series showing the presence of considerable "phase" information.

We divide the series of interest into smaller portions and calculate the appropriate RQA statistics. It is known that for small enough portions, the data should exhibit stationarity. This property is exploited in spectral analysis in construction of time-dependent spectra. The size of the highest window of stationary data is an important characteristic and allows us to detect possible structural breaks. The series analysed here should not show structural breaks, because Kostov (2001) corrected them using unobserved components models. Therefore only non-linear structural shifts could be identified in the data.

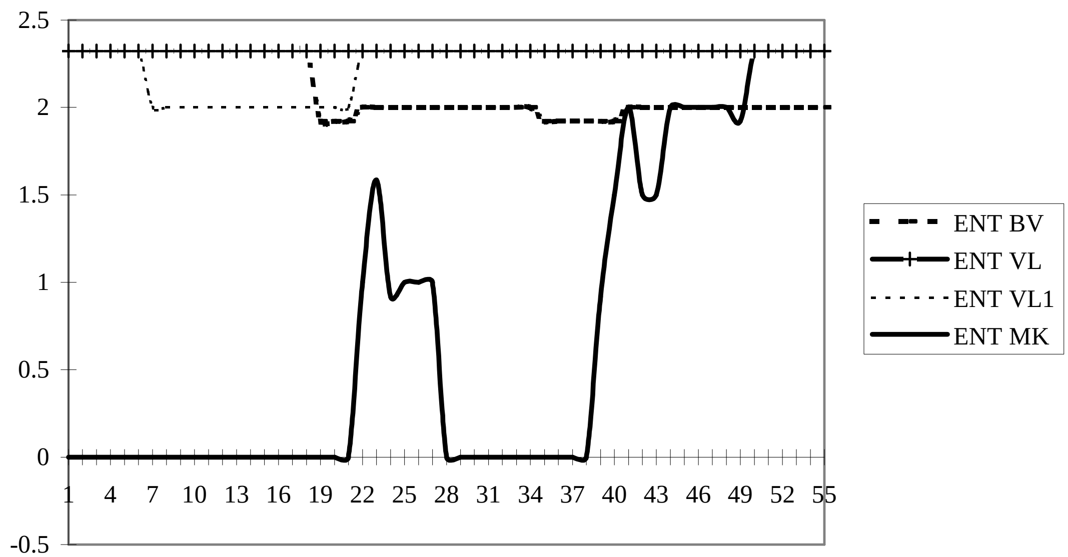
Table 5.1. Estimated values for the maximum recurrent line in the recurrence plots for the differenced series with different sets of parameters

	Beef		Veal		Milk	
Embedding dimension	Radius	Max line	Radius	Max line	Radius	Max line
20	0.5	33	0.28	41	0.58	56
	0.6	62	0.3	41	0.6	56
	0.7	63	0.35	59	0.7	61
	0.8	63	0.4	62	0.8	61
	0.91	64	0.42	65	0.85	63
30	0.67	39	0.35	31	0.65	46
	0.69	40	0.38	46	0.7	46
	0.75	52	0.4	52	0.75	50
	0.8	52	0.45	53	0.8	50
	0.9	53	0.46	54	0.9	53
40	0.8	42	0.4	21	0.75	36
	0.85	42	0.45	39	0.8	36
	0.9	43	0.5	43	0.9	41
	1	43	0.54	44	1	43
	1.1	45				
50	0.85	31	0.45	20	0.9	29
	0.9	32	0.48	28	0.95	31
	0.95	33	0.5	30	1	31
	1	33	0.55	33	1.05	31
	1.1	35	0.57	34	1.1	33
60	0.95	23	0.5	18	1	21
	0.98	23	0.53	22	1.05	21
	1	23	0.56	22	1.1	21
	1.1	23	0.59	23	1.15	21
	1.15	25	0.6	24	1.2	23

Using 20 observations (months) windows we get a stationary beef series for the beginning of the period. The first non-stationary windowed series is the one beginning in February 1990, that is the window February 1990 - October 1991. This is consistent with the structural shift in the period (Feb. 1991- Apr. 1992) reported by Kostov (2001). It appears that the removal of this break is unsatisfactory. The latter however is probably due to the model misspecification (omitted components) that we have discussed in the previous section.

The fundamental changes in the structure of the series can be assessed via the change in the level of entropy presented in graph 1. These show that the index for beef has switched between three different regimes of behaviour. The first could be related to the pre-transition period and one could assume that future behaviour would be determined by the other two regimes. Inspecting the recurrence plot for the beef index, and more specifically the width of the recurrence area around the main diagonal, one can see the switches between different regimes.

Graph 1. Entropy level changes for recurrence relationships



20-observations windows

Surprisingly the index for household consumption of veal exhibits a constant level of entropy. This can be interpreted as a stability in the technology of production and "consumption" of veal within households. The 20-observations windows are however non-stationary in the first part of the period, gradually becoming stationary, beginning with the window (Jan 91 - Sep. 92). The fundamental relation stays the same; this means that the rate of change (volatility) of the veal series has slowed down after 1991 allowing for longer windows of relative stability of the "auto-correlation" function.

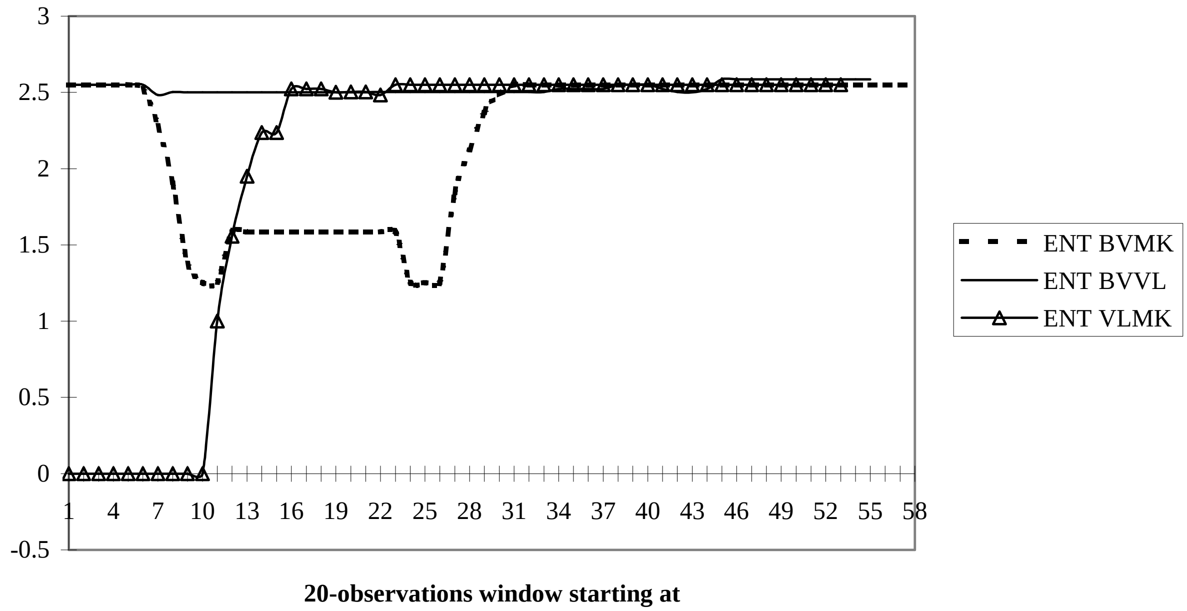
Insofar as the analysis is conditional on the value of the radius, as defined in the recurrence plot, that is our definition of "neighbourhood", we can "zoom in" by slightly decreasing this parameter. This would not change the results for stationary series, but can provide useful details in non-stationary cases, where the internal relationship can vary on different scales. In this case, denoted on graph 1 as VL1 we qualitatively get the same results as before. Due to the changed scale however, the beginning of the stationary windows is displaced later in time and a drop in entropy is found in windows 7 through to 21. This drop can be related to increased macroeconomic instability since early 1991, at beginning of land reform in 1992 allowing for restoration of the original level of entropy. At this lower scale the non-stationary 20-observations windows reappear at the end of the period, starting with the 47th window. The latter demonstrates the stability of the relationship and explains why the picture of transition to a relationship that is closer to stationarity from the greater scale of analysis is not accompanied by a corresponding change in the level of entropy. In this case we have a scale effect.

The 20-observation windows for the milk index reveal interesting behavioural patterns. At the beginning of the period the windowed series look like a random non-deterministic system. A temporary rise in the level of entropy is observed in windows 22 through to 27, which contain price liberalisation and beginning of the land reform. The series are characterised by zero entropy until window 39 when a gradual increase in entropy begins. Since the 50th window, the series become stationary.

We are primarily interested in the joint structure of the series. To investigate it we construct cross-recurrence plots and calculate the appropriate RQA statistics. Whilst univariate recurrence plots are symmetric, the cross-recurrence ones are not usually. Insofar as they relate data points on different time and different variables we can exploit their asymmetry to extract some qualitative information about the nature of the relationship between the variables. After a visual inspection one can see the white left upper and right lower corners of all cross-recurrence plots. This indicates the lack of recurrence relationship between temporaly distant values in the series. In the beef-veal plot there are, on average, more recurrence points in the lower triangle. This means that the influence of the veal series on the beef series is greater and lasts longer than the other way influence. In the middle of the period this relationship is almost one-way. We can define the relationship between two variables, when they are correlated for earlier observations of one variable and later observations of the other, but not vice versa as "causal". Every point in the cross-recurrence plots represents a relationship between a pair of observation of the two series. Therefore we can say that there is a greater "causality" from veal to beef than in the opposite direction. This relationship however undergoes considerable change in early 1995 when it becomes more homogenous in both directions. The latter has significant implications for any attempt to quantitatively model this relationship. Such models should be able to accommodate a structural break at the beginning of 1995 and to relate this structural break to some factors.

The relationship between beef and milk also shows interesting developments. At the beginning of the period one can identify "causality" from milk to beef. This pattern continues, interrupted by intervals of non-regular behaviour until late 1994 - early 1995. Then a structural break takes place that makes the temporal relationship two-way, although now the "causality" from beef to milk is stronger. This structural break seems to take place roughly two or three months before the one in the relationship between beef and veal. The latter is a logical result of the beef chain characteristics, where a decision on changes in future veal production has immediate impact on production and consumption of milk, the effect for veal coming later.

Graph 2. Entropy changes for cross recurrence relationships



This suggests one should expect a similar pattern in the case of veal and milk. This is the case, but the structural break comes significantly earlier - at the end of 1991. In other aspects the relationship between veal and milk is similar to that between beef and milk.

What is the meaning of the structural changes? Transforming the recurrence relationship from one into two way implies increased co-ordination between the dynamics of these variables. The reason for this recurrence is not necessarily causal, but may be due to the coordinated dynamics within the system. One way in which this can be expressed is the econometric notion of co-integration. It is clear that a co-integration relationship that links the above three variables can not exist before 1995. There is no guarantee that one will exist after 1995, but results showing co-integration over the whole period would be spurious. Expressing recurrence relationships in terms of a co-integration relationship is simplistic, because the former is non-linear and can not be compared to the linear notion of co-integration. There will be a direct link between them only in the case of linear time series and relationships. Employing smaller data windows can help us to better situate the changes in the relationship between the pairs of variables. The similar entropy levels at the end of the period confirm our argument about increased "coordination" between the series dynamics.

6. Reinterpreting the results: what is non-stationarity?

Conventional econometric methodology dealing with non-stationary data is built around the assumption for integrated time series with an integer order of integration. This means that by differencing the original data once (for I(1) processes) or more, one can get stationary data. Modern co-integration analysis thus treats 1 (and other integers) as special value(s). This is however a contemporary econometric myth, that is an assumption that is desirable and should be true. Time series with an integer order of integration are a particular case of the much more general class of fractionally integrated processes. In the case of non-integer order of integration, applying differencing leads to yet another fractionally integrated process. The latter may exhibit spurious correlations for virtually the same reasons, as those in the case of spurious regression analysis of I(1) processes (see Phillips, 1986). Modern co-integration theory is valid only if the order of integration is an integer value and does not ensure against spurious regression results which gave rise to its development.

Since the non-integer order of integration of the time series may be a source of the non-stationarity we estimate the fractional integration parameter d for the whole sample and different sub-samples and present the results in table 2. One could object to the validity of this estimation with such a small sample, but we use it here for illustrative purposes. First let us look at the results for the whole sample. A necessary pre-condition for existence of a co-integrating relationship is that the two highest values for the fractional differencing parameter are equal. In this case these are the d-values for beef and veal. Taking into account the interval estimates for d, that is the standard errors, we conclude that they are statistically different and therefore no co-integrating relationship is possible. Owing to the relatively small sample size when we use sub-samples, the standard errors increase and the corresponding intervals begin to overlap. It is necessary to construct reliable statistical test to test for the significance of the hypothesis that the two highest d-values are equal. Its power however will be small in the case of such small samples. When we fix the starting or ending point of the sub-samples and gradually decrease or increase them, one can clearly see the variability in the estimates for the fractional difference parameter. The same phenomenon is observed when comparing the results from consecutive two year periods. This variability is a source of non-stationarity. It indicates a changing over time data generating mechanism. This type of non-stationarity can not be removed by differencing transformations, because the transformed series will exhibit the same properties. The milk index is a good illustration of this point. The estimates for the fractional differencing parameter for the whole sample and the different sub-samples are below 0.5 and one would be tempted to conclude that this series is stationary. This will be the case if the d-values were stable, that is if the data generating mechanism was not changing. This is not the case. RQA parameters show the general non-stationarity in the data. Employing a statistical test based on the stability of the fractional differencing parameter, could help in distinguishing between difference removable and other types of non-stationarity. This could be useful for model identification because it will rule out some proposed models. The latter would require sufficient data length. If this is not the case, the resulting high standard errors in the estimation of the fractional differencing parameter may invalidate the above decomposition.

Table 6.1. Estimated values for the fractional differencing parameter, d.

	d	s.e.	t-value		d	s.e.	t-value
all sample				until 02.1991			
Beef	0.638159	0.07286	8.76	Beef	0.691009	0.28540	2.42
Veal	0.878877	0.08216	10.70	Veal	0.321238	0.25990	1.24
Milk	0.285406	0.07801	3.66	Milk	-0.550633	0.24330	-2.26
Since 1990				until 1992			
Beef	0.633180	0.07847	8.07	Beef	0.483825	0.21230	2.28
Veal	0.868130	0.08075	10.80	Veal	0.761365	0.15250	4.99
Milk	0.285168	0.09187	3.10	Milk	0.206484	0.15590	1.32
since 02.1991				until 1993			
Beef	0.574654	0.08470	6.78	Beef	0.548886	0.13710	4.00
Veal	0.640840	0.07433	8.62	Veal	0.820452	0.11670	7.03
Milk	0.067942	0.10160	0.67	Milk	0.249892	0.12790	0.95
since 1992				until 1994			
Beef	0.483013	0.10480	4.61	Beef	0.601726	0.10130	5.94
Veal	0.706967	0.11300	6.26	Veal	0.815360	0.09940	8.20
Milk	0.057499	0.11500	0.50	Milk	0.270422	0.10300	2.63
since 1993				until 1995			
Beef	0.425320	0.12550	3.37	Beef	0.624996	0.08323	7.51
Veal	0.720424	0.13440	5.36	Veal	0.835052	0.09218	90.60
Milk	-0.401745	0.13310	-3.02	Milk	0.273154	0.08773	3.11
since 1994				02.1991 - 12.1992			
Beef	-0.340594	0.21810	-1.56	Beef	0.413895	0.18870	2.19
Veal	0.954730	0.15850	6.02	Veal	0.517595	0.16120	3.21
Milk	-0.522430	0.15600	-3.35	Milk	0.067869	0.25440	0.27
Since 1995				01.1993 - 12.1994			
Beef	0.023214	0.27670	0.08	Beef	0.421701	0.16910	2.49
Veal	0.555347	0.33220	1.67	Veal	0.099900	0.26140	0.38
Milk	-1.255050	0.21920	-5.73	Milk	-0.514149	0.19460	-2.64

The stability of the co-integrating space has been investigated in Hansen and Johansen (1999). Similar test procedures may be applied in the case of fractional co-integration. An interesting option would be to estimate the models in a state space form when allowing for variable parameters. The variance of the latter would provide an estimate of the degree of non-stationarity. Unfortunately it is not clear how to present a fractionally differencing process in a state space form. An option may be to estimate the order of the process and difference it and then to formulate a space state model for the transformed series, thus relying on indirect estimates for the stability of the fractional differencing parameter.

The analysis clearly demonstrates that non-stationarity is a much wider concept than that of integration. Moreover, while there are tools and methods for dealing with non-linear data, they still require stationarity which is difficult to obtain. Elaborating methods for better understanding this problem are therefore highly desirable. Once again we return to the recurrence plots of the differenced series presented in appendix 2. It appears that parts of these recurrence plots exhibit patterns that are typical for quasi-periodic, while other parts look typical for chaotic data. It is difficult to make definite statements on the basis of such a small sample, but it looks like the effect of non-stationarity, that is of changing system behaviour can be expressed in forcing the system into a sequence of qualitatively different states of motion.

7. Conclusions

RQA provides powerful tools for analysis of both stationary and non-stationary, linear and non-linear data. We emphasise the non-stationarity and non-linearity properties. These are areas where this tool can bring advantages to the analysis. We have chosen data that is known to exhibit features such as non-linear structural breaks in order to demonstrate the usefulness of the recurrence analysis. It has been shown that when the data is intrinsically non-linear, linear and conditionally linear models can not adequately capture the underlying dynamics.

It has to be stressed that the recurrence analysis tools presented are data exploratory rather than model building tools. They can not replace econometric models. The investigation into the nature of the economic data, however, has important implications for empirical modelling. When the data exhibits the properties of a random system, as in the case of the milk index in the beginning of the period, conventional regression type models may be misleading. The variable character of the relationships, revealed in our analysis needs further clarification. One has to identify the likely reasons for the detected structural changes. This is equivalent to endogenising these changes within the modelling framework.

The variable nature of these relationships however does not allow for blind application of regression-like models. The use of some regime switching type of models for predicting such relationships is also of restricted use. They are known to have poor predicting performance and one can not be sure about the date and magnitude of future breaks. Therefore the only alternative is to apply models with local characteristics, such as unobserved components models with local trend and level, being aware that local quantities are the best predictors one can obtain in the short and medium terms, although they would have poor performance in the long term. The stochastic trend that these models employ is the price we pay for our inability to properly understand and model the changes that cause the varying data generation mechanism, that is, for the unknown source of non-stationarity. In practical terms this is expressed in the restricted forecasting abilities of our models. Identification of the likely determinants of such a non-linear and variable process that allows better assessment of its long term features is difficult. It would require vast amounts of good quality data and does not guarantee that the long-term future of the system will be genuinely predictable. When the sources of this volatility are omissions and inadequacies in our theoretical models however, we can improve on them.

One of the sources of non-stationarity is the change in systems parameters. Such changes are usually defined exogenously in conventional economic models. We can illustrate the above point by the example of preferences.
It is well known that preferences cannot be considered as fixed over time as neo-classical economic theory postulates. If this was the case, the consumption patterns would have been stationary. Strotz (1956) demonstrated that in a dynamic context, the stationarity of consumption which is necessary for the consistency of preferences is often violated. Thaler (1981) presents further empirical evidence. The latter inconsistency, however can be accommodated as a process of interaction between conflicting short and long term preferences (Hoch and Loewenstein, 1991). To put it simply, the structural break in preferences that leads to non-stationarity in consumption can in principle be endogenised. Nevertheless this does not guarantee that the problem of non-stationarity would be overcome. Non-stationarity seems to be a property of time and it may probably persist in economic time series.

References:

Beaulieu, J.J. and J.A. Miron (1993) Seasonal Unit Roots in Aggregate US Data, *Journal of Econometrics*, **55**, 305-328.

Casdagli, M.C. (1997) Recurrence plots revisited, *Physica D*, **108(1-2)**, 12-44.

Eckmann, J.P., S.O. Kamphorst and D. Ruelle (1987) Recurrence plots of dynamic systems, *Europhysiscs Letters*, **4(9)**, 973-977.

Hansen, H. and S. Johansen (1999) Some Tests for Parameter Constancy in Cointegrated VAR models, *Econometrics Journal*, **2(2)**, 306-333.

Harrison, J. and C. F. Stevens (1976). Bayesian Forecasting (with discussion). *Journal of the Royal Statistical Society*, **Series B 38**, 205–47.

Harvey, A. and S.J. Koopman (2000) Signal Extraction and the Formulation Of Unobserved Component Models, *Econometrics Journal*, **3(1)**, 84-107.

Harvey, A. C., S. J. Koopman, and J. Penzer (1998) "Messy time series" In T. *Advances in Econometrics*, vol. 13, Eds. B. Fomby and R. C. Hill, New York: JAI Press.

Hegger, R., H. Kantz and T. Schreiber (1999) Practical implementation of non-linear time series methods: the TISEAN package, *Chaos*, **9(2)**, 413-435.

Hoch, S.J. and G.F. Loewenstein (1991) Time-inconsistent Preferences and Consumer Self-control, *Journal of Consumer Research*, **17**, 492-507.

Koopman, S.J, N. Shepherd and J.A. Doornik (1999) Statistical algorithms for models in state space using SsfPack 2.2**,** *Econometrics Journal*, **2**, 113ñ166.

Kostov, P. (2001) Assessing and measuring the importance of small-scale subsistence farming in Bulgarian agriculture, Unpublished PhD thesis, Dept. of Agricultural Economics and Food Marketing, University of Newcastle.

Lancaster, K.J. (1966) A new approach to consumer theory, *Journal of Political Economy*, **74**, 132-157.

Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics, *Journal of Econometrics*, **33**, 311-340.

Strotz, R.H. (1956) Myopia and Inconsistency in Dynamic Utility Maximisation, *Review of Economic Studies*, **23**, 165-180.

Thaler, R.H (1981) Some Empirical Evidence on Dynamic Inconsistency, *Economic Letters*, **8**, 201-207.

Urbach , R.M.A. (2000) *Footprints of chaos in the markets : analyzing non-linear time series in financial markets and other real systems*, Financial Times, Prentice Hall, London.

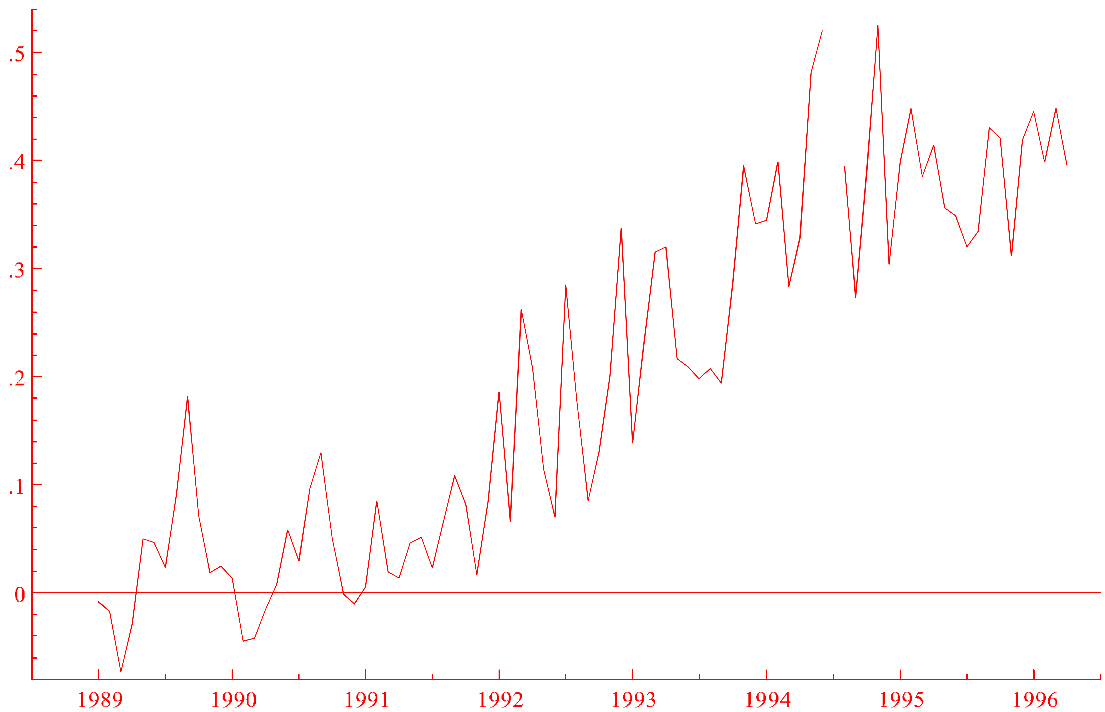
Webber, C.L.and J.P. Zbilut (1994) Dynamical assessment of psychological systems and states using recurrence plot strategies, *Journal of Applied Physiology*, **76**, 965-973.

Yu, D.J., W.P. Lu and R.G. Harison (1999) Detecting dynamical nonstationarity in time series data, *Chaos*, **9(4)**, 867-870.

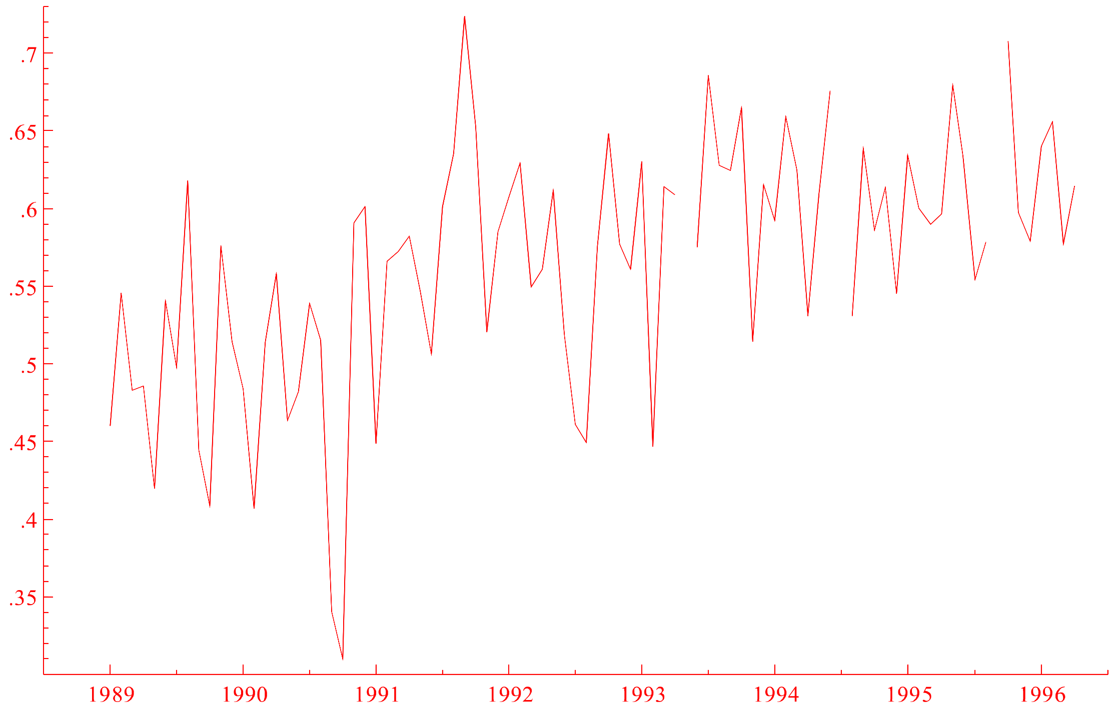
Zbilut, J.P. and C.L. Webber (1992) Embeddings and delays as derived from quantification of recurrence plots, *Physics Letters A*, **171**, 191-203.

Appendix 1. Subsistence indices for beef, milk and veal

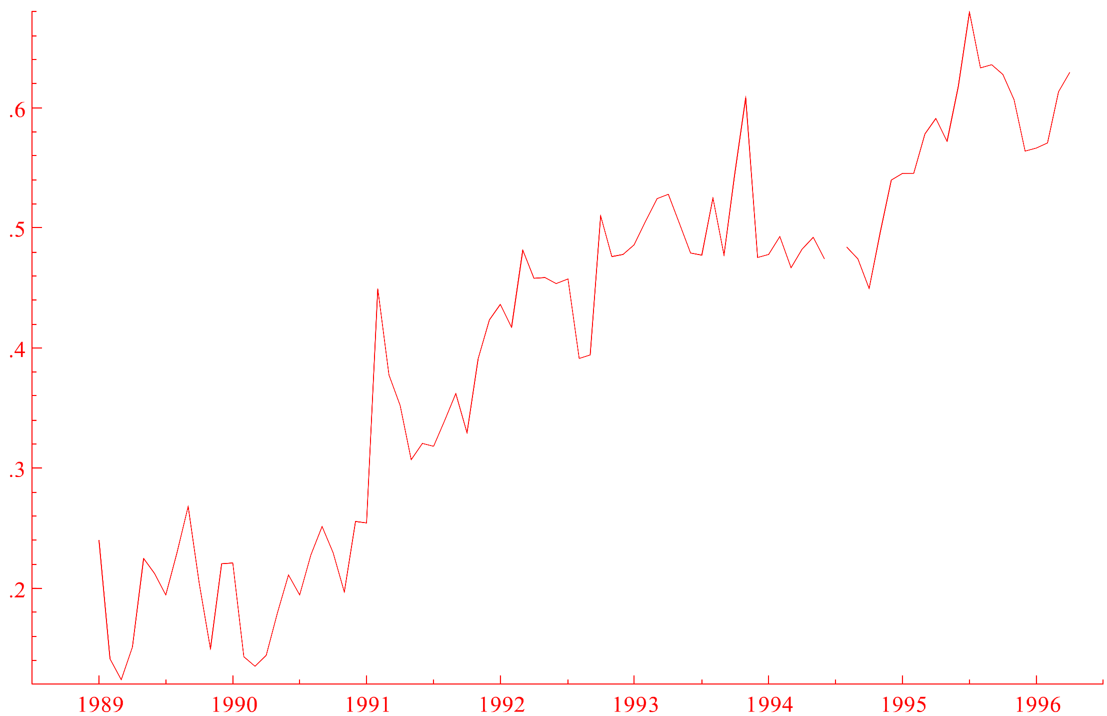
Subsistence index for beef



Subsistence index for milk

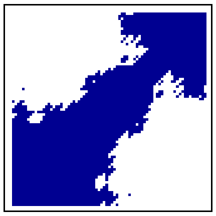


Subsistence index for veal

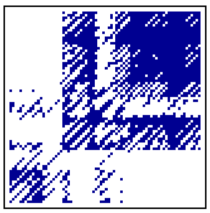


Appendix 2. Recurrence and cross-recurrence plots

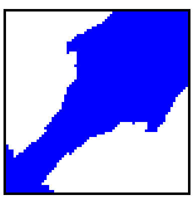
Beef



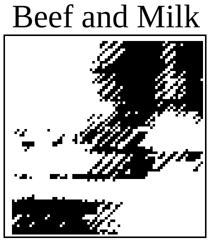
Milk



Veal

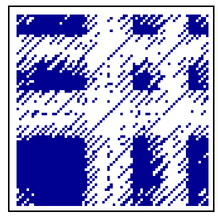


Beef and Milk

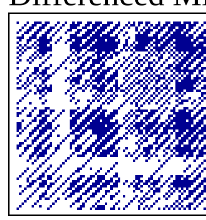


Veal and Milk

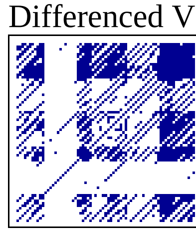
Differenced Beef



Differenced Milk



Differenced Veal



Beef and Veal